Please cite this article as:

Heyvaert, M., Maes, B., Van den Noortgate, W., Kuppens, S., & Onghena, P. (2012). A multilevel meta-analysis of single-case and small-n research on interventions for reducing challenging behavior in persons with intellectual disabilities. *Research in Developmental Disabilities, 33,* 766-780. doi:10.1016/j.ridd.2011.10.010

A multilevel meta-analysis of single-case and small-*n* research on interventions for reducing challenging behavior in persons with intellectual disabilities.

Heyvaert, M., Maes, B., Van den Noortgate, W., Kuppens, S., & Onghena, P.

Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven, Belgium

Correspondence concerning this article can be addressed to Mieke Heyvaert, Methodology of Educational Sciences Research Group, Andreas Vesaliusstraat 2 - Box 3762, B-3000 Leuven, Belgium. Phone +32 16 326265. Fax +32 16 326200.

E-mail Mieke.Heyvaert@ppw.kuleuven.be

Abstract

The effectiveness of different interventions for challenging behavior (CB) in persons with intellectual disabilities (ID) was reviewed by means of a two-phase study. First, a systematic review of 137 meta-analyses and reviews on group-study interventions for CB in persons with ID was conducted. Based on this review, hypotheses concerning the effectiveness of divergent interventions for CB and concerning the impact of variables moderating treatment effectiveness were systematically generated. Second, these hypotheses were tested by means of a multilevel meta-analysis of single-case and small-n research. Two hundred and eighty-five studies reporting on 598 individuals were examined. The average treatment effect was large and statistically significant. However, this effect varied significantly over the included studies and participants. Compared to the meta-analyses and reviews focusing on group-studies in this research domain, the results of the present multilevel meta-analysis of single-case and small-n intervention research provided more detailed knowledge on which specific CB and intervention components moderate the interventions' effectiveness.

Keywords: meta-analysis; single-case; review; challenging behavior; problem behavior; intellectual disability; mental retardation

Research highlights

Effectiveness articles of interventions on challenging behavior (CB) in persons with ID are studied. ► Single-case data are combined using a hierarchical linear model meta-analysis. ► A large and statistically significant overall treatment effect is found, but the effect varies over participants and studies. ► CB topographies and intervention components significantly moderate treatment effects.

A multilevel meta-analysis of single-case and small-*n* research on interventions for reducing challenging behavior in persons with intellectual disabilities

1. Introduction

Individuals with intellectual disabilities (ID) are particularly vulnerable for developing challenging behaviors (CBs) (Došen et al., 2007; Emerson et al., 2001; Kahng et al., 2002; McIntyre et al., 2002). Those CBs place the individual and his family at risk of serious harm to their physical and psychological well-being and they can prevent them from having ordinary experiences in the community (Beck et al., 2004; Benson & Brooks, 2008; Blacher & McIntyre, 2006; Cooper et al., 2009; Didden et al., 2006; Gavidia-Payne & Hudson, 2002; Harvey et al., 2009; Hassiotis et al., 2008; Hastings, 2002; Knapp et al., 2005; Matson & Boisjoli, 2009; McIntyre et al., 2002; McMillan et al., 2004). Nowadays, divergent intervention strategies are used to reduce CB among persons with ID, including biological, psychotherapeutic and contextual intervention components (Antonacci et al., 2008; Beail, 2003, Deb et al., 2007; Došen & Day, 2001; Gavidia-Payne & Hudson, 2002; Grey & Hastings, 2005; Heyvaert et al., 2010; Kahng et al., 2002; McGillivray & McCabe, 2006; Prout & Nowak-Drabik, 2003; Sohanpal et al., 2007; Sturmey, 2004; Taylor, 2002, 2005; Willner, 2005).

Answering the call of Campbell (2003) in this journal to systematically compare behavioral and pharmacological treatment outcomes for reducing CB by means of a quantitative review, we include the broad range of pharmacological, behavioral and contextual intervention studies in the present study. By doing that, we intend to address the question that remained unanswered by previously published SCS*n* reviews in this research domain (e.g., Campbell, 2003; Denis et al., 2011; Didden et al., 2006; Hart & Banda, 2010; Harvey et al., 2009; Shogren et al., 2004): do substantial effectiveness differences exist for

different contextual, behavioral, and pharmacological treatments for CB among persons with ID?

Many primary-level studies have been published in the domain of intervention research for reducing CB among persons with ID, and a considerable amount of these studies concern small numbers of participants. Meta-level research is needed to examine and compare the broad range of intervention components studied in those primary-level articles: combining the treatment results from several participants studied under diverging circumstances provides insights concerning the general applicability of treatment-effect findings and concerning factors that moderate the treatment effect (Van den Noortgate & Onghena, 2008).

Although several recent articles provide a review on interventions for CB for persons with ID (e.g., Brylewski & Duggan, 2004; Chan et al., 2010; Dinca et al., 2005; Hassiotis & Hall, 2008; Heyvaert et al., 2010; Hogg et al., 2001; Lotan & Gold, 2009; Matson & Neal, 2009; Parikh et al., 2008; Prout & Nowak-Drabik, 2003; Shogren et al., 2004; Sohanpal et al., 2007; Taylor, 2002; Thomson et al., 2009a, 2009b), with only a few exceptions (e.g., Campbell, 2003; Denis et al., 2011; Didden et al., 2006; Hart & Banda, 2010; Harvey et al., 2009; Shogren et al., 2004) these meta-analyses and reviews focus on group-studies, and exclude a sizeable part of the empirical evidence that is published in this research domain: the single-case and small-*n* (SCS*n*) research. In contrast to group-studies that investigate the effectiveness of treatments by comparing groups and present *aggregated* data for a *group* of participants, SCS*n* studies report effectiveness data for each participant *separately*. SCS*n* designs refer to experiments in which one participant (SC) or a small number of participants (S*n*) is repeatedly observed under the levels of at least one manipulated independent variable (Onghena, 2005). Compared to the group-comparison design, some advantages of the SCS*n* design in the research domain of interventions for reducing CB in persons with ID are its focus on the individual providing an in-depth insight into the behavior of a single case, the

allowance of detailed analysis of characteristics of ‘nonresponders’ as well as ‘responders’ to interventions, the study of behavior evolution through a large number of repeated observations, and its cost-effective approach (Horner et al., 2005; Van den Noortgate & Onghena, 2007).

It is suggested that the results of published SCS*n* research could significantly differ from group-studies (e.g., Newcombe, 1987; Thornton & Lee, 2000). Accordingly, concerning a single research topic a meta-analysis exclusively including SCS*n* research could lead to different conclusions than a meta-analysis exclusively including group-studies. By systematically comparing the results of recently published SCS*n* research with the conclusions of meta-analyses and systematic reviews based on group-studies on interventions for reducing CB in persons with ID, the present study intends to examine whether there exists substantial differences between both groups of studies.

In order to achieve this goal, a meta-analysis of SCS*n* research will be performed, preceded by a systematic review of meta-analyses and reviews on group-study interventions for CB in persons with ID. The preceding systematic review will provide an overview of the recent developments in this domain, and will allow to systematically generate hypotheses concerning the effectiveness of divergent interventions for CB, and concerning the impact of variables moderating this treatment effectiveness. These hypotheses will be tested in the meta-analysis of SCS*n* research on this topic. Finally, we will answer the posed question whether there exists substantial differences between recently published SCS*n* research, and meta-analyses and systematic reviews based on group-studies in the research domain of interventions for reducing CB in persons with ID. This approach is unprecedented in this research domain.

Supplementary to the published SCS*n* meta-analyses on treatments of CBs in persons with ID that mainly apply nonregression approaches like percentages of nonoverlapping data (PND)[1] (Campbell, 2003; Didden et al., 2006; Hart & Banda, 2010; Harvey et al., 2009; Shogren et al., 2004), we choose to use a regression approach for quantifying outcomes of single-subject research. We applied an adaptation of the general hierarchical linear model that can be used for a multilevel meta-analysis of SCS*n* research measures of effect (Bryk & Raudenbush, 1992; Van den Noortgate & Onghena, 2003a, 2003b), in which the variance in observed treatment effects is split up in sampling variance, variance between participants from the same study, and variance between studies, and in which we try to explain this variation by the inclusion of case and study characteristics. Advantages of this approach are described in papers of Van den Noortgate and Onghena (2003c, 2007, 2008).

Summarizing, the present study intends to answer the following questions: (a) What is the overall effect of different contextual, behavioral, and pharmacological interventions for CB among persons with ID?; (b) Which characteristics of participants, of the context of the study, and of the intervention moderate this effect?; and (c) Are the conclusions for SCS*n* studies and group-studies equivalent?

2. Generating hypotheses concerning moderating variables

2.1. Literature search and criteria for inclusion

Relevant reviews were identified through systematic searches of three electronic databases as well as hand searches of 32 key scientific journals known to publish research on

[1] PND summarizes the efficacy of single-subject interventions by calculating the percentage of intervention data points that do not overlap with the highest or lowest baseline data points (Campbell, 2003). A recent quantitative synthesis of Maggin et al. (2011) showed that PND is the most often used effect-size metric in the meta-analysis of single-subject research for students with disabilities.

treatments for persons with ID. Afterwards, we systematically examined the bibliographies of the identified articles.

We included reviews describing intervention effects for CB among persons with ID, and possibly also reporting on variables moderating the intervention effects. All types of quantitative and qualitative reviews were eligible (e.g., statistical meta-analysis, narrative review, meta-study, meta-synthesis, meta-summary, aggregated analysis). However, we excluded reviews exclusively based on SCS*n* research (e.g., Campbell, 2003; Denis et al., 2011; Didden et al., 2006; Hart & Banda, 2010; Harvey et al., 2009; Shogren et al., 2004), in order to avoid overlap between our two databases: the meta-analyses and reviews database, and the SCS*n* database. We only included reviews published in the period January 2000 - April 2011. We did not apply language-related inclusion/exclusion criteria. We did not assign weights to the retrieved articles, e.g., based on the type of the review.

First, the electronic searches of the databases *Eric*, *Pubmed*, and *Web of Science* involved the combining of search strings for CB with search strings for ID (Table 1), and with the search terms *'review'*, *'synthesis',* and *'meta-analysis'*. This search located 103 reviews that met our inclusion criteria. Second, a manual search of reviews published in the 32 journals summed up in Table 2 was conducted, by systematically screening the indexed articles. This search *additionally* retrieved 30 reviews that met our inclusion criteria. Third, the bibliographical lists of all reviews obtained through the electronic databases and journals searches were scrutinized. Nine reviews not yet identified were retrieved. In order to avoid overlap between the two databases, we did not include six reviews based on SCS*n* research. The final database included 136 reviews, all written in English.

We systematically generated an overview of the main results of the retrieved reviews in Microsoft Excel, and cross-compared and tabulated the variables reported to have a possible moderating effect on the intervention effects.

2.2. Hypotheses and study variables

Based on the systematic review of the meta-analyses and reviews on group-study interventions for CB in persons with ID, we hypothesized that the interventions would on average lead to a reduction of CB, since the greater part of the meta-analyses and reviews concluded that the studied interventions led to a reduction of CB. Concerning pharmacological interventions, some authors claim that there is convincing evidence for the effectiveness of some medication - particularly for the atypical antipsychotic medication risperdone - (e.g., Barnard et al., 2002; Deb et al., 2007; Grey & Hastings, 2005), although many authors assert that pharmacological interventions for the present lack sufficient empirical effectiveness and should not be the first treatment option (e.g., Antonacci et al., 2008; Benson & Brooks, 2008; Brylewski & Duggan, 2004; Dinca et al., 2005; La Malfa et al., 2006; Matson & Neal, 2009; Sohanpal et al., 2007; Taylor, 2002; Thomson et al., 2009a, 2009b). Additionally, concerns are raised about their adverse effects (e.g., Aman et al., 2000; Berry-Kravis & Potanos, 2004; Deb et al., 2007; La Malfa et al., 2006; Matson et al., 2000; Parikh et al., 2008; Sohanpal et al., 2007). Regarding psychotherapeutic and contextual interventions, there exist several studies advocating their effectiveness (e.g., Ager & O'May, 2001; Carr et al., 2009; Gavidia-Payne & Hudson, 2002; Grey & Hastings, 2005; Harvey et al., 2009; Prout & Nowak-Drabik, 2003; Shogren et al., 2004), although some studies conclude that the evidence is still too scant (e.g., Antonacci et al., 2008; Gustafsson et al., 2009; Hassiotis & Hall, 2008; Sturmey, 2004; Willner, 2005).

Additionally, based on the review of the retrieved meta-analyses and reviews we systematically generated a list of variables that were reported to have a possible moderating impact on the treatment effectiveness (Table 3). Based on our review, we hypothesized that especially the target CB, the specific intervention components (bio/psycho/social), and the presence of pretreatment functional analysis could function as moderators.

3. Methods

3.1. Literature search and criteria for inclusion

Parallel to the literature search for meta-analyses and reviews concerning group-studies, for the multilevel meta-analysis of SCS*n* relevant primary-level research reports published in the period January 2000 - April 2011 were identified through systematic searches of (1) the databases *Eric*, *Pubmed*, and *Web of Science* applying the search strings shown in Table 1, (2) 31 key scientific journals summed up in Table 2 (the journal *Developmental Disabilities Research Reviews* was excluded from this search, since this journal only publishes reviews, and not primary-level SCS*n* studies), and (3) the bibliographical lists of the papers identified in steps (1) and (2).

The literature search for this meta-analysis aimed at retrieving *single-case studies* and *small-n studies* reporting on intervention effects on CB among persons with ID. This contrasts with the data collection for the qualitative preparatory phase, that pursued to retrieve *reviews* including group-studies on this topic. Research reports that met the following criteria were included: (a) The study aim was the reduction of CB among persons with ID; (b) The single-case or small-*n* study reported measures of the CB of a participant under baseline and treatment conditions, with each containing at least two data points. A study was excluded if a group of participants was described but no individual characteristics were presented or if the raw data representing the level of CB under baseline and treatment conditions was not

reported for each participant separately or could not be accurately retrieved from for instance a graph. Consequently, studies only reporting percentages of reduction of CB were excluded, as well as studies only reporting aggregated data for multiple participants. We included all types of SCS*n* designs that answered to our inclusion/exclusion criteria. Again, we did not apply language-related inclusion/exclusion criteria.

First, the electronic searches of the three databases located 172 unique research reports that met our inclusion criteria. Second, the manual search of papers published in the 31 journals additionally retrieved 108 research reports that met our inclusion criteria. Third, searching the bibliographical lists of all included papers identified five other relevant studies. The final database included 285 studies reporting on 598 individuals. All included studies were written in English. There were 155 single-case studies and 130 *small n (* > 1) studies.

3.2. Data extraction

3.2.1. Study variables

All retrieved studies were reviewed and coded for each of the variables described in Table 3 and Table 4. Several participant, context, and intervention variables were coded, as well as the meta-analytic variables *Publication year* and *Study quality*. In order to code the latter variable we applied the Single-case Experimental Design (SCED) Scale developed by Tate et al. (2008): ten items were used to assess the methodological quality of and the use of statistical analyses in each included SCS*n* study. The SCED Scale is characterized by high levels of inter-rater reliability. In the study of Tate et al. (2008) inter-rater reliability was first studied between six experienced raters using intra-class correlation (ICC), each rating twenty papers published in the 10-year period 1996 to 2005 that were randomly selected from 312 reports archived in PsycBITE. Inter-rater reliability for the total score was high (for individual raters: overall ICC = .84 with 95% confidence interval .73-.92; for consensus ratings between

pairs of raters: overall ICC = .88 with 95% confidence interval .78-.95). Second, the results were replicated with two independent novice raters who were trained in the use of the scale (ICC = .88 with 95% confidence interval .73-.95).

3.2.2. Treatment effectiveness

Next to the coding of possible moderating variables, information concerning treatment effectiveness was extracted from the research reports. These effectiveness data were presented graphically in the primary studies. The raw data were retrieved with the aid of the UnGraph Version 5 software, a digitizer program that gives X,Y coordinates of lines or points on scanned images (Biosoft, 1997-2011). In a study of Shadish et al. (2009) extraction with UnGraph proved highly valid and reliable over several different kinds of analyses. Afterwards, the raw data were imported in Microsoft Excel.

3.2.3. Reliability of coding

Ten percent of the included studies were randomly selected, and analyzed by two researchers. Interrater agreement was calculated on each study variable by dividing the number of agreements by the number of agreements plus disagreements. The interrater reliability was 92.01%. Disagreements were afterwards resolved by discussion between the two researchers, and the corrected codes were used for the meta-analysis.

3.3. Data analysis

The data of our meta-analysis show a hierarchical three-level structure: 285 primary studies are included in the meta-analysis (between-studies level), that describe 598 individuals with ID and CB (between-participants level), and for each individual repeated measurements of CB are reported (within-participants level). Accordingly, we choose to apply an adaptation of the general hierarchical linear model that can be used for a multilevel meta-analysis of

SCS*n* measures of effect, described in publications of Van den Noortgate and Onghena (2003a, 2003b, 2003c, 2007, 2008). An advantage of this approach is that by modeling the variation within participants, between participants of the same study, and between studies, it accounts for the possible dependency that may result from this three-level nesting (Van den Noortgate & Onghena, 2008). In this model, the coefficient that indicates the magnitude of the effect of the intervention on the CB of a participant can be considered as a standardized difference between means, since it is equal to the difference in condition means, divided by the within-condition standard deviation (Van den Noortgate and Onghena, 2008). However, Van den Noortgate and Onghena (2008) note that standardized mean differences from group-comparison and SC designs are not directly comparable: in SC studies effect sizes are calculated on the basis of scores *from the same participant*, while in group-comparison studies two independent groups are compared and scores within each condition are derived *from different participants*. The standard deviation in a group-comparison study is influenced by between-participant as well as within-participant variation, while the within-phase standard deviation of the scores from a single participant in a SC design refers only to differences within participants (Van den Noortgate and Onghena, 2008).

The SAS software Version 9.2 was used to conduct the present multilevel meta-analysis of SCS*n* studies on interventions for reducing CB in persons with ID. The SAS procedure MIXED provided estimates and tests of the overall effect, the overall intercept, and the covariance parameters (see Table 5). Descriptive statistics for each variable (see Table 4) and tests on the moderating impact of each variable on the treatment effectiveness were conducted (see Table 5).

Additionally, using the metafor package in R (Viechtbauer, 2010) we checked for publication bias by generating a funnel plot (Rothstein et al., 2005) and testing for funnel plot asymmetry (Egger et al., 1997a).

4. Results

4.1. Three-level model

First, let us look at the three-level random effects regression model without moderators, presented as 'Model 1' in Table 5 (see Van den Noortgate & Onghena, 2008). This analysis reveals that on average the interventions for CB are highly effective: in comparison to the baseline conditions, the level of CB is 2.96 standard deviations lower in the treatment conditions. A Wald test acknowledges that this reduction in CB is statistically significant, Z = -13.16, $p < .0001$. Looking at the covariance parameter estimates, we notice that the intervention effects vary significantly over studies, with an estimated variance of 3.32, Z = 3.67, $p = .0001$. The intervention effects vary significantly over the included participants too, with an estimated variance of 20.57, Z = 15.01, $p < .0001$.

Second, in order to examine which variables can explain this variation of the intervention effects over the participants and the studies, we look at the three-level meta-analysis model including all potential moderators (see Van den Noortgate & Onghena, 2008) that are listed in Tables 3 and 4. In table 5, Model 2 shows a statistically significant moderating effect of the variables *Age* (Z = -2.80, $p = .0051$), *Diagnosis of autism spectrum disorder* (Z = -3.43, $p = .0006$), the CB-type *Aggression* (Z = 10.29, $p < .0001$), the CB-type *Destructive behavior* (Z = 2.59, $p = .0096$), the intervention component *Manipulating antecedent factors* (Z = -2.80, $p = .0052$), and the intervention component *Informing, educating, training the environment* (Z = -2.67, $p = .0076$). For the other variables, no statistically significant moderating effect is found. However, testing a meta-analytical model including a large amount of potential moderators brings along the risk of misinterpretation due to multicollinearity. Therefore, we added each of the potential moderators separately to the regression model. This analysis confirms that there is a statistically significant moderator

effect of these six variables: *Age* (Z = -9.94, $p < .0001$), *Diagnosis of autism spectrum disorder* (Z = -6.25, $p < .0001$), the CB-type *Aggression* (Z = 6.09, $p < .0001$), the CB-type *Destructive behavior* (Z = -3.23, $p = .0013$), the intervention component *Manipulating antecedent factors* (Z = -5.29, $p < .0001$), and the intervention component *Informing, educating, training the environment* (Z = -6.83, $p < .0001$). Interpreting these results for Model 2, we see that the interventions for CB on average turn out to be less effective for persons with aggression or destruction as CB-type, and more effective for included participants with a diagnosis of autism spectrum disorder. When participants are treated by an intervention including the component *Manipulating antecedent factors* or *Informing, educating, training the environment*, the results are on average better. The expected effect for persons treated by an intervention at older age is on average better compared to the effects for younger participants[2].

Third, we combined these six moderators in the final three-level meta-analysis model ('Model 3' in Table 5), excluding other predictors (see Van den Noortgate & Onghena, 2008). Model 3 shows a statistically significant moderating effect of the variables *Age* (Z = -2.15, $p = .0312$), *Diagnosis of autism spectrum disorder* (Z = -2.91, $p = .0037$), the CB-type *Aggression* (Z = 9.76, $p < .0001$), and the intervention component *Manipulating antecedent factors* (Z = -2.42, $p = .0155$). However, in contrast to Model 2, the variables *Destructive behavior* and *Informing, educating, training the environment* do not play a significant moderating effect in Model 3 (respectively Z = 1.77, $p = .0762$; Z = -1.74, $p = .0810$). The interpretation of the results for Model 3 is analogue to Model 2: interventions on average turn out to be less effective for persons with aggression as CB-type. Interventions are on average more effective for persons with a diagnosis of autism spectrum disorder, for older participants, and when the intervention includes the component *Manipulating antecedent factors*.

[2] For the continuous moderator *Age*, the regression coefficient (in Model 2 this coefficient is -0.06) indicates the increase of the effect when the age increases with one unit (i.e., one year).

4.2. Sensitivity analysis

Figure 1 displays the box and whisker diagram of the standardized random effects for the individual participants. There seem to be six outlying participants (all belonging to different studies) deviating more than three standard deviations from zero, including one extremely outlying case showing a standardized random effect of 15.64. Since these outlying effects do not seem to be based upon incorrect values (that could lead to a distorted view of the population), the participants are not omitted from the analysis. However, as a check of the influence of these outliers on the conclusions, we performed a sensitivity analysis by comparing the full dataset, the dataset without the one extreme outlier, and the dataset without the six outliers. By removing these outliers, the overall effect of -2.96 (which means that the level of CB is on average 2.96 standard deviations lower in the treatment conditions, compared to baseline conditions) is reduced to respectively -2.90 and -2.69. For all three datasets, the Wald test shows that on average this reduction in CB is statistically significant, with respectively: Z = -13.16, $p < .0001$; Z = -14.58, $p < .0001$; and Z = -18.44, $p < .0001$. This means that for all three datasets the interventions are on average highly effective in reducing CB.

Applying the analyses described in section 4.1. under 'Model 1' for the databases without the one extreme outlier, and without the six outliers, we see that the variance of effect between participants for the latter database is considerably lower compared to the original database. However, the intervention effects still vary significantly over the included participants for both databases: the estimated variance is 11.76 for the database without the one extreme outlier (SD = 0.95; Z = 12.36, $p < .0001$) and 3.48 for the database without the six outliers (SD = 0.30; Z = 11.54, $p < .0001$). So, even after omitting the six most extreme scoring participants from the analysis, the intervention effects still vary significantly over the included participants.

When we implement the moderator analyses (described in section 4.1. under 'Model 2') for the database without the six outliers, we conclude that from the six variables that had a significant moderating effect for the original database (i.e., *Age*, *Diagnosis of autism spectrum disorder*, *Aggression*, *Destructive behavior*, *Manipulating antecedent factors*, and *Informing, educating, training the environment*), only four variables remain to have a significant moderating effect: *Diagnosis of autism spectrum disorder* ($Z = -2.17$, $p = .0299$), the CB-types *Aggression* ($Z = 9.95$, $p < .0001$) and *Destructive behavior* ($Z = 2.23$, $p = .0258$), and the intervention component *Manipulating antecedent factors* ($Z = -2.70$, $p = .0070$). Interpreting these results for Model 2 for the database without the six outliers, we see that the interventions for CB on average turn out to be less effective for persons with aggression or destruction as CB-type, and more effective for included participants with a diagnosis of autism spectrum disorder. When participants are treated by an intervention including the component *Manipulating antecedent factors* the intervention effects are on average better.

Subsequently, we combined these four moderators in the final three-level meta-analysis model for the database without the six outliers, excluding other predictors ('Model 3'). Model 3 only shows a statistically significant moderating effect of the variables *Aggression* ($Z = 9.41$, $p < .0001$) and *Manipulating antecedent factors* ($Z = -2.94$, $p = .0033$). In contrast to Model 2, the variables *Destructive behavior* and *Diagnosis of autism spectrum disorder* do not play a significant moderating effect in Model 3 (respectively $Z = 1.51$, $p = .1306$; $Z = -1.35$, $p = .1781$). The interpretation of these final results are that interventions on average turn out to be less effective for persons with aggression as CB-type; and that they are on average more effective when the intervention includes the component *Manipulating antecedent factors*.

Summarizing the moderator analysis results for the original database, of all tested variables six where found to have a statistically significant moderating effect (for 'Model 2':

Age; *Aggression; Destructive behavior; Diagnosis of autism spectrum disorder; Informing, educating, training the environment;* and *Manipulating antecedent factors*). When subsequently only these six variables were combined in one model, only four of them remained to show statistically significant moderating effects (for 'Model 3': *Age, Aggression, Diagnosis of autism spectrum disorder,* and *Manipulating antecedent factors*). However, when omitting the six most extreme participants from the moderator analysis (see Figure 1), of all tested variables four where found to have a statistically significant moderating effect (for 'Model 2': *Aggression, Destructive behavior, Diagnosis of autism spectrum disorder,* and *Manipulating antecedent factors*). When subsequently only these four variables were combined in one model, only two of them remained to show statistically significant moderating effects (for 'Model 3': *Aggression* and *Manipulating antecedent factors*).

4.3. Publication bias analysis

Each black dot in Figure 2 represents the observed outcome for one participant (horizontal axis) plotted against the corresponding standard error (vertical axis). The vertical line in Figure 2 shows the estimate based on Model 1, the random effects model without moderators. A pseudo confidence interval area is depicted around this estimate, with bounds equal to +-1.96 times the standard error value (Viechtbauer, 2010). As expected, by visually inspecting this plot we find some evidence for publication bias: the plot is not symmetric. On the left side of the plot, we notice six outlying points, referring to six outlying participants that were also detected in Figure 1. The regression test for funnel plot asymmetry with standard error as predictor described by Viechtbauer (2010) confirms that there is significant asymmetry in the funnel plot: $t(596) = -7.03$, $p < .0001$.

4.4. Comparing the effects for SCSn- and group-studies

In the following paragraph, we discuss the comparison between the hypotheses based on the studied meta-analyses and systematic reviews, and the results of our multilevel meta-analysis of SCS*n* research concerning (a) the effectiveness of biological, psychotherapeutic and contextual interventions for CB among persons with ID, and concerning (b) the impact of variables moderating the treatment effectiveness.

(a) The generated hypothesis stating that the interventions would on average lead to a reduction of CB (see section 2.2.) was confirmed: our meta-analysis of SCS*n* research yields a high overall intervention effect that is statistically significant. Looking at Table 4, we find evidence endorsing the positive effects of psychotherapeutic and contextual intervention components, but no evidence for an overall positive effect of pharmacological interventions. These findings correspond to what we concluded from the preceding systematic review of meta-analyses and reviews based on group-study interventions for CB in persons with ID (see section 2.2.).

(b) We hypothesized that all variables listed in Table 3 could have a moderating impact on the treatment effectiveness, but that especially the target CB, the specific intervention components (bio/psycho/social), and the presence of pretreatment functional analysis would function as variables moderating the treatment effectiveness (see section 2.2.). The hypotheses suggesting the potential influence of the target CB (for the variable *Aggression* - the other moderating CB effects are not significant) and the specific intervention components (for the variable *Manipulating antecedent factors* - the other moderating intervention component effects are not significant) were confirmed in the meta-analysis of SCS*n* studies. Inconsistent with our hypothesis, we found no evidence that the intervention effects significantly depend on the presence of pretreatment functional analysis: looking at

Table 4 we see that the treatment effects for studies including pretreatment functional analysis are in general better than in studies lacking pretreatment functional analysis, but this effect is not significant (Table 5). After controlling for six extreme outliers, our multilevel meta-analysis of SCS*n* research found no significant evidence for the moderating effect of other variables besides the variables *Aggression* and *Manipulating antecedent factors* (see section 4.2.).

Concluding from (a) and (b), we state the results of the multilevel meta-analysis of SCS*n* research to a large extent correspond with the conclusions of meta-analyses and reviews focusing on group-studies in this research domain; and that the present multilevel meta-analysis provides more detailed knowledge on which specific CB (i.e., *Aggression*) and intervention components (i.e., *Manipulating antecedent factors*) significantly moderate the overall intervention effect.

5. Discussion

The present multilevel meta-analysis aimed to systematically study and compare the effectiveness of different interventions for CB in people with ID that are described in SCS*n* research, and to examine the impact of variables moderating the treatment effectiveness. Since the overall intervention effect is high (-2.96) and statistically significant ($Z = -13.16$, $p < .0001$), we conclude that on average the interventions for CB in people with ID reported in the SCS*n* studies are highly effective.

However, we have to make three comments concerning the generalization of our results. First, although the overall intervention effect is high, the effect varies significantly over studies (the estimated variance is 3.32, $Z = 3.67$, $p = .0001$) and over the included participants (the estimated variance is 20.57, $Z = 15.01$, $p < .0001$). For the variance of effect between studies, this result means that based on the normal distribution, we can say that 95%

of the study-effects vary from -6.53 (i.e., -2.96 – 1.96*$\sqrt{3.32}$) to + 0.61 (i.e., -2.96 + 1.96*$\sqrt{3.32}$), with negative values indicating reductions in CB (desired), and positive values indicating increases in CB. Likewise, for the variance of effect between participants, this result means that based on the normal distribution, we expect for a typical study that 95% of the participants-effects vary from -11.85 (i.e., -2.96 – 1.96*$\sqrt{20.57}$) to + 5.93 (i.e., -2.96 + 1.96*$\sqrt{20.57}$). We see that especially the variance of effect between participants is extremely high, pointing to the fact that it is possible that the interventions do not have the desired effect for several participants. Looking at the datasets without the one extreme outlier and without the six outliers (see 4.2.), the variance of effect between participants is considerably lower compared to the original database, although the intervention effects still vary significantly over the included participants for both databases: the estimated variance is respectively 11.76 (SD = 0.95; Z = 12.36, $p < .0001$) and 3.48 (SD = 0.30; Z = 11.54, $p < .0001$). So, although even after omitting the six most extreme scoring participants from the analysis the Wald test showed that on average the reduction in CB is still statistically significant (see section 4.2.), these intervention effects vary significantly over the included participants.

A second comment concerns publication bias. After visually inspecting the generated funnel plot (Figure 2), the regression test for funnel plot asymmetry confirmed the significant asymmetry of this plot and the presence of publication bias (see 4.3.). Publication bias boils down to the problem that since 'negative' (i.e., finding no reduction of the CB due to the intervention, but an increase of the CB) and null-finding (i.e., finding nor a reduction nor an increase of the CB due to the intervention) studies are often not reported and published, and therefore not included in reviews and meta-analyses, most treatments tend to be less effective in clinical practice than the published research suggests. In comparison to group-studies, especially SCS*n* research is more unlikely to be reported and published unless describing significant results (Thornton & Lee, 2000). Additionally, although we did not apply language-

related inclusion/exclusion criteria, all studies retrieved by the electronic databases search and the bibliographical lists search were written in English. Our search in relevant journals only concerned English-written journals. The fact that all included studies are written in English could additionally explain some of the found publication bias: authors are more likely to publish their studies in an English-language journal if the results are statistically significant (Egger et al., 1997b; Rothstein et al., 2005).

Third, there was some overlap between the primary-level research studies included in the meta-analyses and reviews, that were analyzed in our systematic review preceding our multilevel meta-analysis. Some of the primary-level research studies were discussed in more than one meta-analysis or review included in our systematic review. Consequently, it is possible that some of the primary-level research studies had a relatively stronger influence on the conclusions of our systematic review than other primary-level research studies. Since several of the meta-analyses and reviews analyzed in our systematic review did not include a list describing all included primary-level research studies, it was difficult to address this drawback.

Notwithstanding these limitations, the present study has several strengths. First, it is one of the few meta-analyses in the domain of intervention research on CB in people with ID that includes SCS*n* research. Although SCS*n* research represents a vast part of the empirical evidence published in this research domain, it is often omitted from meta-analyses. Solely relying on a certain part of the available evidence (i.e. group-studies) in meta-analyses and reviews can endanger the correctness of the generalizations made to the entire population of persons with ID and CB. One explanation for the absence of SCS*n* research in most meta-analyses is that although a single-case (SC) design is considered as a suitable design for drawing valid conclusions about one individual, it is statistically incorrect to generalize the results of a SC study to a whole population. However, exploring the generalizability of SC

results is feasible by replicating the results over other cases, or by aggregating the results from several single cases in a meta-analysis (Van den Noortgate & Onghena, 2007, 2008). Another explanation is that most of the methodological literature on meta-analysis techniques has focused on group-studies (Van den Noortgate & Onghena, 2008). However, during the last decades the question how to quantify results from SC studies for purposes of quantitative review is more often brought up for discussion (e.g., Allison & Gorman, 1993; Campbell, 2004; Maggin et al., 2011; Van den Noortgate & Onghena, 2008; White et al., 1989). The present meta-analysis illustrates that combining a large number of studies each concerning only one or a few participants through a variant of the general hierarchical linear model can lead to insightful and important results.

Second, the present study is the only one in this research domain that systematically compares the results of recently published SCS*n* research with the results of meta-analyses and systematic reviews based on group-studies. In section 4.4., we concluded that the results of the multilevel meta-analysis of SCS*n* intervention research on CB for persons with ID to a large extent correspond with the results of meta-analyses and reviews focusing on group-studies in this research domain. However, the present multilevel meta-analysis of SCS*n* research provides more detailed knowledge on which specific CB and intervention components significantly moderate the overall intervention effect.

Furthermore, the found large average treatment effect is consistent with the effects reported in previous SCS*n* meta-analyses on interventions for CB in persons with ID (Campbell, 2003; Denis et al., 2011; Didden et al., 2006; Hart & Banda, 2010; Harvey et al., 2009; Shogren et al., 2004). Variables that were found to have a significant moderator effect in these meta-analyses are: presence of pretreatment functional assessment (Campbell, 2003; Didden et al., 2006; Harvey et al., 2009), sensory impairment (Denis et al., 2011), study design (Didden et al., 2006), duration of treatment (Harvey et al., 2009), diagnosis of autism

(Harvey et al., 2009), and gender (Shogren et al., 2004). Notice that only the variable *presence of pretreatment functional assessment* was found to be a significant moderator in more than one of these SCS*n* meta-analyses. Although all of these moderators were also included in the present SCS*n* meta-analysis, after the excluding of the six most extreme outliers we did not found significant moderating effects for these variables.

Third, we mention the extensive moderator analyses as a final strength of the present study. Based on the preceding review of meta-analyses and reviews of group-studies in this research domain, we systematically generated a list of all variables that were reported to have a possible moderating impact on the treatment effectiveness, including twelve variables related to participant characteristics, ten variables concerning the context of the study, fourteen intervention-related variables, and two meta-analytic variables (Table 3). We coded all 285 included studies for these 38 variables, computed descriptive statistics and effect sizes and associated standard errors for each moderator (Table 4), and generated the three-level meta-analysis model including the potential moderators (Table 5). Two variables turned out to have a significant moderating impact on the treatment effectiveness, even after controlling for six extreme outliers (see section 4.2.): *Aggression* and *Manipulating antecedent factors*.

Our results indicate that the interventions for CB on average turn out to be significantly less effective for participants with outwardly directed aggression as CB-type, when compared to treated participants without outwardly directed aggression as CB-type. Consistent with our findings, the SCS*n* meta-analysis of Didden et al. (2006) concerning the behavioral treatment of CB in individuals with mild ID found relative high mean effect sizes associated with stereotypic behavior, and the lowest mean effect sizes associated with destructive behaviors and verbal aggression. Additionally, Harvey et al. (2009) found in their SCS*n* meta-analysis of interventions for CB in children with ID that participants with disruptive and aggressive behavior generally responded least well to behavior change efforts,

while children with self-injury, stereotypy, socially inappropriate, and destructive behavior responded best to the interventions.

Finally, we conclude based on our results that when participants are treated by an intervention including the component *Manipulating antecedent factors*, the results are on average better than when an intervention without one of these components used. Manipulating antecedent factors that increase the probability that a person shows CB has proven to be an effective intervention strategy in the ID literature (e.g., Brosnan & Healy, 2011; Conroy & Stichter, 2003; Krantz et al., 1993). Often-researched types of antecedent factors in CB studies are environmental, instructional, psychological, and social factors (Conroy & Stichter, 2003). For example, interventions such as reducing task difficulty, making the daily schedule more predictable, and providing choice-making opportunities can rather easily be implemented, and can strongly decrease the occurrence of CB.

To conclude, we summarize the answers provided to the three posed research questions. (a) The overall effect of different contextual, behavioral, and pharmacological interventions for CB among persons with ID was large and statistically significant. However, this effect varied significantly over the included studies and participants. (b) From the 38 coded characteristics of participants, of the context of the study, and of the intervention, after controlling for six extreme outliers, significant evidence was only found for the moderator effect of the CB topography *aggression*, and the intervention component *manipulating antecedent factors*. (c) The results of the multilevel meta-analysis of SCS*n* research to a large extent correspond with the conclusions of meta-analyses and reviews focusing on group-studies in this research domain. However, the present multilevel meta-analysis provides more detailed knowledge on which specific CB and intervention components significantly moderate the overall intervention effect. Additionally, the found large average treatment effect is consistent with the effects reported in previous SCS*n* meta-analyses on interventions for CB

in persons with ID. On a regular base, future research should update the present findings, in order to add to the evidence base, and to allow researchers and practitioners to make grounded decisions when aiming to reduce CB in persons with ID.

Acknowledgements

Mieke Heyvaert is a predoctoral fellow of the Research Foundation - Flanders (Belgium).

Sofie Kuppens is a postdoctoral fellow of the Research Foundation - Flanders (Belgium).

References

Ager, A., & O'May, F. (2001). Issues in the definition and implementation of "best practice" for staff delivery of interventions for challenging behavior. *Journal of Intellectual and Developmental Disability, 26*, 243–256.

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavioral Research Therapy, 31*, 621–631.

Aman, M. G., Collier-Crespin, A., & Lindsay, R. L. (2000). Pharmacotherapy of disorders in mental retardation. *European Child & Adolescent Psychiatry, 9*, 98–107.

Antonacci, D. J., Manuel, C., & Davis, E. (2008). Diagnosis and treatment of aggression in individuals with developmental disabilities. *Psychiatry Quarterly, 79*, 225–247.

Barnard, L., Young, A. H., Pearson, J., Geddes, J., & O'Brien, G. (2002). A systematic review of the use of atypical antipsychotics in autism. *Journal of Psychopharmacology, 16*, 93–101.

Beail, N. (2003). What works for people with mental retardation? Critical commentary on cognitive-behavioural and psychodynamic psychotherapy research. *Mental Retardation, 41*, 468–472.

Beck, A., Hastings, R. P., Daley, D., & Stevenson, J. (2004). Pro-social behaviour and behaviour problems independently predict maternal stress. *Journal of Intellectual & Developmental Disability, 29*, 339–349.

Benson, B. A., & Brooks, W. T. (2008). Aggressive challenging behaviour and intellectual disability. *Current Opinion in Psychiatry, 21*, 454–458.

Berry-Kravis, E., & Potanos, K. (2004). Psychopharmacology in Fragile X Syndrome--Present and future. *Mental Retardation and Developmental Disabilities Research Reviews, 10*, 42–48.

Biosoft (1997-2011). *UnGraph Version 5* [Computer software]. Retrieved from http://www.biosoft.com/w/ungraph.htm

Blacher, J., & McIntyre, L. L. (2006). Syndrome specificity and behavioural disorders in young adults with intellectual disability: Cultural differences in family impact. *Journal of Intellectual Disability Research, 50*, 184–198.

Brosnan, J., & Healy, O. (2011). A review of behavioral interventions for the treatment of aggression in individuals with developmental disabilities. *Research in Developmental Disabilities, 32*, 437–446.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA, US: Sage Publications.

Brylewski, J., & Duggan, L. (2004). Antipsychotic medication for challenging behaviour in people with learning disability. *Cochrane Database of Systematic Reviews*, CD000377.

Campbell, J. M. (2003). Efficacy of behavioural interventions for reducing problem behaviour in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities, 24*, 120–138.

Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification, 28*, 234–246.

Carr, J. E., Severtson, J. M., & Lepper, T. L. (2009). Noncontingent reinforcement is an empirically supported treatment for problem behavior exhibited by individuals with developmental disabilities. *Research in Developmental Disabilities, 30*, 44–57.

Chan, S. W., Thompson, D. R., Chau, J. P., Tam, W. W., Chiu, I. W., & Lo, S. H. (2010).The effects of multisensory therapy on behaviour of adult clients with developmental disabilities: A systematic review. *International Journal of Nursing Studies, 47*,108–122.

Conroy, M. A. & Stichter, J. P. (2003). The application of antecedents in the functional assessment process - Existing research, issues, and recommendations. *The Journal of Special Education, 37*, 15–25.

Cooper, S. A., Smiley, E., Allan, L. M., Jackson, A., Finlayson, J., Mantry, D., & Morrison, J. (2009). Adults with intellectual disabilities: Prevalence, incidence and remission of self-injurious behaviour, and related factors. *Journal of Intellectual Disability Research, 53*, 200–216.

Deb, S., Sohanpal, S. K., Soni, R., Len-o Tre, L., & Unwin, G. (2007). The effectiveness of antipsychotic medication in the management of behaviour problems in adults with intellectual disabilities. *Journal of Intellectual Disability Research, 51*, 766–777.

Denis, J., Van Den Noortgate, W., & Maes, B. (2011). Self-injurious behavior in people with profound intellectual disabilities: A meta-analysis of single-case studies. *Research in Developmental Disabilities, 32*, 911–923.

Didden, R., Korzilius, H., van Oorsouw, W., & Sturmey, P. (2006). Behavioural treatment of challenging behaviours in individuals with mild mental retardation: Meta-analysis of single-subject research. *American Journal on Mental Retardation, 111*, 290–298.

Dinca, O., Paul, M., & Spencer, N. J. (2005). Systematic review of randomized controlled trials of atypical antipsychotics and selective serotonin reuptake inhibitors for behavioural problems associated with pervasive developmental disorders. *Journal of Psychopharmacology, 19*, 521–532.

Došen, A., & Day, K. (2001). *Treating mental illness and behaviour disorders in children and adults with mental retardation*. Washington: American Psychiatric Publications.

Došen, A., Gardner, W. I., Griffiths, D. M., King, R., & Lapointe, A. (2007). *Practice guidelines and principles: Assessment, diagnosis, treatment and related support services for persons with intellectual disabilities and problem behaviour.* Netherlands: Centre of Consultation and Expertise.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997a). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

Egger, M., Zellweger-Zähner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997b). Language bias in randomised controlled trials published in English and German. *Lancet, 350*, 326–329.

Emerson, E., Kiernan, C., Alborz, A., Reeves, D., Mason, H., Swarbrick, R., et al. (2001). The prevalence of challenging behaviours: A total population study. *Research in Developmental Disabilities, 22*, 77–93.

Gavidia-Payne, S., & Hudson, A. (2002). Behavioural supports for parents of children with an intellectual disability and problem behaviours: An overview of the literature. *Journal of Intellectual & Developmental Disability, 27,* 31–55.

Grey, I. M., & Hastings, R. P. (2005). Evidence-based practices in intellectual disability and behaviour disorders. *Current Opinion in Psychiatry, 18,* 469–475.

Gustafsson, C., Ojehagen, A., Hansson, L., Sandlund, M., Nystrom, M., Glad, J., et al. (2009). Effects of psychosocial interventions for people with intellectual disabilities and mental health problems: A survey of systematic reviews. *Research on Social Work Practice, 19,* 281–290.

Hart, S. L., & Banda, D. R. (2010). Picture exchange communication system with individuals with developmental disabilities: A meta-analysis of single subject studies. *Remedial and Special Education, 31,* 476–488.

Harvey, S. T., Boer, D., Meyer, L. H., & Evans, I. M. (2009). Updating a meta-analysis of intervention research with challenging behaviour: Treatment validity and standards of practice. *Journal of Intellectual & Developmental Disability, 34,* 67–80.

Hassiotis, A., & Hall, I. (2008). Behavioural and cognitive-behavioural interventions for outwardly-directed aggressive behaviour in people with learning disabilities. *Cochrane Database of Systematic Reviews*, CD003406.

Hassiotis, A., Parkes, C., Jones, L., Fitzgerald, B., & Romeo, R. (2008). Individual characteristics and service expenditure on challenging behaviour for adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 21,* 438–445.

Hastings, R. P. (2002). Parental stress and behaviour problems of children with developmental disability. *Journal of Intellectual & Developmental Disability, 27,* 149–160.

Heyvaert, M. , Maes, B., & Onghena, P. (2010). A meta-analysis of intervention effects on challenging behaviour among persons with intellectual disabilities. *Journal of Intellectual Disability Research, 54,* 634–649.

Hogg, J., Cavet, J., Lambe, L., & Smeddle, M. (2001). The use of 'Snoezelen' as multisensory stimulation with people with intellectual disabilities: A review of the research. *Research in Developmental Disabilities, 22,* 353–372.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Kahng, S., Iwata, B. A., & Lewin, A. B. (2002). Behavioural treatment of self-injury, 1964 to 2000. *American Journal on Mental Retardation, 107,* 212–221.

Knapp, M., Comas-Herrera, A., Astin, J., Beecham, J., & Pendaries, C. (2005). Intellectual disability, challenging behaviour and cost in care accommodation: What are the links? *Health and Social Care in the Community, 13,* 297–306.

Krantz, P. L., MacDuff, M. T., & McClannahan, L. E. (1993). Programming participation in family activities for children with autism. Parents' use of photographic activity schedules. *Journal of Applied Behavior Analysis, 26,* 137–138.

La Malfa, G., Lassi, S., Bertelli, M., & Castellani, A. (2006). Reviewing the use of antipsychotic drugs in people with intellectual disability. *Human Psychopharmacology: Clinical and Experimental, 21,* 73–89.

Lotan, M., & Gold, C. (2009). Meta-analysis of the effectiveness of individual intervention in the controlled multisensory environment (snoezelen) for individuals with intellectual disability. *Journal of Intellectual & Developmental Disability, 34,* 207–215.

Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985-2009. *Exceptionality, 19,* 109–135.

Matson, J. L., Bamburg, J. W., Mayville, E. A., Pinkston, J., Bielecki, J., Kuhn, D., et al. (2000). Psychopharmacology and mental retardation: A 10 year review (1990- 1999). *Research in Developmental Disabilities, 21,* 263–296.

Matson, J. L., & Boisjoli, J. A. (2009). Restraint procedures and challenging behaviours in intellectual disability: An analysis of causative factors. *Journal of Applied Research in Intellectual Disabilities, 22,* 111–117.

Matson, J. L., & Neal, D. (2009). Psychotropic medication use for challenging behaviors in persons with intellectual disabilities: An overview. *Research in Developmental Disabilities, 30,* 527–586.

McGillivray, J. A., & McCabe, M. P. (2006). Emerging trends in the use of drugs to manage the challenging behaviour of people with intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 19,* 163–172.

McIntyre, L. L., Blacher, J., & Baker, B. L. (2002). Behaviour/mental health problems in young adults with intellectual disability: The impact on families. *Journal of Intellectual Disability Research, 46,* 239–249.

McMillan, D., Hastings, R. P., & Coldwell, J. (2004). Clinical and actuarial prediction of physical violence in a forensic intellectual disability hospital: A longitudinal study. *Journal of Applied Research in Intellectual Disabilities, 17,* 255–265.

Newcombe, R. G. (1987). Towards a reduction in publication bias. *British Medical Journal, 295,* 656–659.

Onghena, P. (2005). Single-case designs. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science*, vol.4 (pp. 1850–1854). Chichester: Wiley.

Parikh, M. S., Kolevzon, A., & Hollander, E. (2008). Psychopharmacology of aggression in children and adolescents with autism: A critical review of efficacy and tolerability. *Journal of Childhood and Adolescent Pharmacology, 18,* 157–178.

Prout, H. T., & Nowak-Drabik, K. M. (2003). Psychotherapy with persons who have mental retardation: An evaluation of effectiveness. *American Journal on Mental Retardation, 108,* 82–93.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. John Wiley & Sons, Chichester, England.

Shadish, W. R., Brasil, I. C. C., llingworth, D. A. I., White, K. D., Galindo, R., Nagler, E. D., & Rindsk, D. M. (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. *Behavior Research Methods, 41,* 177–183.

Shogren, K. A., Faggella-Luby, M. N., Bae, S. J., & Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior: A meta-analysis. *Journal of Positive Behavior Interventions, 6,* 228–237.

Sohanpal, S. K., Deb, S., Thomas, C., Soni, R., Lenôtre, L., & Unwin, G. (2007). The effectiveness of antidepressant medication in the management of behaviour problems in adults with intellectual disabilities: A systematic review. *Journal of Intellectual Disability Research, 51,* 750–765.

Sturmey, P. (2004). Cognitive therapy with people with intellectual disabilities: A selective review and critique. *Clinical Psychology and Psychotherapy, 11,* 222–232.

Tate, R. L., McDonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation, 18,* 385–401.

Taylor, J. L. (2002). A review of the assessment and treatment of anger and aggression in offenders with intellectual disability. *Journal of Intellectual Disability Research, 46,* 57–73.

Taylor, J. L. (2005). In support of psychotherapy for people who have mental retardation. *Mental Retardation, 43,* 450–453.

Thomson, A., Maltezos, S., Paliokosta, E., & Xenitidis, K. (2009a). Amfetamine for attention-deficit hyperactivity disorder in people with intellectual disabilities. *Cochrane Database of Systematic Reviews, 1,* CD007009.

Thomson, A., Maltezos, S., Paliokosta, E., & Xenitidis, K. (2009b). Risperdone for attention-deficit hyperactivity disorder in people with intellectual disabilities. *Cochrane Database of Systematic Reviews, 2,* CD007011.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53,* 207–216.

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18,* 325–346.

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers, 35,* 1–10.

Van den Noortgate, W., & Onghena, P. (2003c). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement, 63,* 765–790.

Van den Noortgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst Today, 8,* 196–209.

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention, 2,* 142–151.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48.

White, D. M., Rusch, F. R., Kazdin, A. E., & Hartmann, D. P. (1989). Applications of meta-analysis in individual subject research. *Behavioral Assessment, 11,* 281–296.

Willner, P. (2005). The effectiveness of psychotherapeutic interventions for people with learning disabilities: A critical overview. *Journal of Intellectual Disability Research, 49,* 73–85.

Table 1

Search strings for "challenging behavior" and "intellectual disability", combined with a Boolean AND (between both search strings), and Boolean OR's (within both search strings).

Search strings for challenging behavior	Search strings for intellectual disability
challenging behavio*	intellectual disabilit*
aberrant behavio*	mental retardation
maladaptive behavio*	developmental disabilit*
problem behavio*	learning disabilit*
behavio* problems	intellectual impairment
self-injury	mental handicap
self-harm	mental deficiency
self-injurious behavio*	
stereotypy	
stereotyped behavio*	
repetitive behavio*	
aggression	
destructive behavio*	
property destruction	
disruptive behavio*	

Table 2
Manually searched journals.

American Journal on Intellectual and Developmental Disabilities (American Journal on Mental Retardation)	International Journal of Disability, Development and Education
Behavior Modification	Journal of Abnormal Child Psychology
Behavior Therapy	Journal of Applied Behavior Analysis
Behavioral Disorders	Journal of Applied Research in Intellectual Disabilities
Behavioral Interventions	Journal of Autism and Developmental Disorders
Behavior Research and Therapy	Journal of Behavior Therapy and Experimental Psychiatry
Brain and Development	Journal of Clinical Child & Adolescent Psychology
British Journal of Clinical Psychology	Journal of Consulting and Clinical Psychology
British Journal of Learning Disabilities	Journal of Developmental and Physical Disabilities
Child and Family Behavior Therapy	Journal of Experimental Child Psychology
Clinical Case Studies	Journal of Intellectual and Developmental Disability
Cognitive and Behavioral Practice	Journal of Intellectual Disability Research
Developmental Disabilities Research Reviews	Journal of Positive Behavior Interventions
Disability & Rehabilitation	Journal of Special Education
Exceptional Children	Research in Developmental Disabilities
Intellectual and Developmental Disabilities (Mental Retardation)	The Journal of ECT

Table 3
Study variables.

Participant variables	Age (in years)
	Gender
	Target CB - Self-injurious behavior
	Target CB - Stereotyped behavior
	Target CB - Aggression
	Target CB - Destructive behavior
	Target CB - Disruptive behavior
	Diagnosis of autism spectrum disorder
	Intellectual disability level [a]
	Sensory impairment
	Motor impairment
	Communicative impairment
Context variables	Design
	Pretreatment functional analysis
	Primary treatment setting
	Primary intervention agent
	Duration of treatment
	Intervention format
	Family involved in treatment
	Peer(s) involved in treatment
	Presence of data/researcher triangulation
	Presence of follow-up data
Intervention variables	Uni- vs. multicomponent intervention
	Biological intervention
	Biological intervention - Medication
	Psychological intervention
	Psychological intervention - Teaching alternative replacement skills
	Psychological intervention - Reward and praise/attention for positive behaviors
	Psychological intervention - Punishment
	Psychological intervention - Use of restraints
	Psychological intervention - Manipulating antecedent factors
	Psychological intervention - Extinction
	Psychological intervention - Gain insight in perception of CB
	Social-contextual intervention
	Social-contextual intervention - Informing, educating, training the environment of the participant (parents, caregivers, staff,…)
	Social-contextual intervention - Adapt the environment to the participant's needs
Meta-analytic variable	Publication year
	Study quality

Notes: CB = challenging behavior; [a] If only a range of intellectual disability level was reported, it was categorized as the lower level.

Table 4

Overview and descriptive statistics for potential moderating variables.

Variable	Value	Descriptive statistics	Moderator effect (SE)
Age	Continuous	M = 18; SD = 14.01; range = 1-65	-0.022 (0.017)
Gender	0 = male; 1 = female	n_0 = 400; n_1= 193	ES_0 -2.842 (0.266); ES_1 -3.194(0.372)
CB type – Self-injurious behavior	0 = no self-injurious behavior; 1 = self-injurious behavior	n_0 = 353; n_1= 245	ES_0 -2.937 (0.232); ES_1 -3.015 (0.242)
CB type – Stereotyped behavior	0 = no stereotyped behavior; 1 = stereotyped behavior	n_0 = 496; n_1= 102	ES_0 -2.913 (0.228); ES_1 -3.211 (0.290)
CB type – Aggression	0 = no aggression; 1 = aggression	n_0 = 282; n_1= 316	ES_0 -3.326 (0.231); ES_1 -2.577 (0.232)
CB type – Destructive behavior	0 = no destructive behavior; 1 = destructive behavior	n_0 = 496; n_1= 102	ES_0 -2.916 (0.226); ES_1 -3.188 (0.244)
CB type – Disruptive behavior	0 = no disruptive behavior; 1 = disruptive behavior	n_0 = 363; n_1= 235	ES_0 -2.886 (0.229); ES_1 -3.091 (0.237)
Diagnosis of autism spectrum disorder	0 = no autism spectrum disorder; 1 = autism spectrum disorder	n_0 = 363; n_1= 235	ES_0 -2.831 (0.260); ES_1 -3.151 (0.291)
Intellectual disability level [a]	1 = borderline; 2 = mild; 3 = moderate; 4 = severe; 5 = profound	n_1 = 20; n_2= 57; n_3 = 86; n_4= 114; n_5 = 109	ES_1 -2.507 (1.288); ES_2 -2.482 (0.748); ES_3 -3.723 (0.585); ES_4 -2.551 (0.512); ES_5 -2.763 (0.538)
Sensory impairment	0 = no sensory impairment; 1 = sensory impairment	n_0 = 561; n_1= 37	ES_0 -2.929 (0.232); ES_1 -3.417 (0.803)
Motor impairment	0 = no motor impairment; 1 = motor impairment	n_0 = 539; n_1= 59	ES_0 -2.986(0.237); ES_1 -2.798 (0.645)
Communicative impairment	0 = no communicative impairment; 1 = communicative impairment	n_0 = 158; n_1= 440	ES_0 -2.965 (0.437); ES_1 -2.955 (0.255)
Design	1 = AB-design; 2 = reversal design; 3 = multiple baseline design; 4 = alternating treatments design	n_1 = 126; n_2= 257; n_3 = 145; n_4= 70	ES_1 -2.187 (0.499); ES_2 -2.710 (0.322); ES_3 -4.218 (0.456); ES_4 -2.682 (0.612)
Pretreatment functional analysis	0 = no pretreatment functional analysis; 1 = pretreatment functional analysis	n_0 = 173; n_1= 425	ES_0 -2.293 (0.439); ES_1 -3.156 (0.260)
Primary treatment setting	1 = community environment; 2 = home; 3 = school; 4 = treatment facility	n_1 = 15; n_2= 71; n_3 = 131; n_4= 374	ES_1 -3.067 (1.413); ES_2 -2.711 (0.298); ES_3 -3.184 (0.341); ES_4 -2.915 (0.254)
Primary intervention agent	1 = investigator; 2 = parent; 3 = peer; 4 = staff; 5 = teacher; 6 = therapist	n_1 = 58; n_2= 64; n_3 = 3; n_4= 93; n_5 = 47; n_6 = 327	ES_1 -2.885 (0.722); ES_2 -2.055 (0.695); ES_3 -4.358 (3.241); ES_4 -4.074 (0.646); ES_5 -2.921 (0.775); ES_6 -2.881 (0.295)
Duration of treatment	1 = 1–5 weeks; 2 = 6–10 weeks; 3 = 11–15 weeks; 4 = 16–20 weeks; 5 = more than 20 weeks; 6 = not reported	n_1 = 32; n_2= 26; n_3 = 11; n_4= 4; n_5 = 134; n_6 = 391	ES_1 -2.221 (0.951); ES_2 -2.980 (1.070); ES_3 -2.741 (1.615); ES_4 -0.758 (2.665); ES_5 -2.605 (0.493); ES_6 -3.158 (0.276)
Intervention format	0 = one-to-one; 1 = group intervention	n_0 = 568; n_1= 30	ES_0 -3.012 (0.229); ES_1 -1.359 (1.093)
Family involved in treatment	0 = no family involved; 1 = family involved	n_0 = 496; n_1= 102	ES_0 -3.093 (0.246); ES_1 -2.332 (0.539)
Peer(s) involved in treatment	0 = no peer(s) involved; 1 = peer(s) involved	n_0 = 559; n_1= 39	ES_0 -2.985 (0.232); ES_1 -2.340 (0.941)
Presence of data/researcher triangulation	0 = no triangulation; 1 = triangulation	n_0 = 41; n_1= 557	ES_0 -2.552 (0.839); ES_1 -2.997 (0.234)
Presence of follow-up data	0 = no follow-up; 1 = follow-up	n_0 = 399; n_1= 199	ES_0 -2.980 (0.269); ES_1 -2.920 (0.397)
Uni- vs. multicomponent intervention	0 = unicomponent intervention; 1 = multicomponent intervention	n_0 = 137; n_1= 461	ES_0 -2.193 (0.483); ES_1 -3.165 (0.251)
Biological intervention	0 = no bio-intervention component included; 1 = bio-intervention component included	n_0 = 520; n_1= 78	ES_0 -3.112 (0.237); ES_1 -1.724 (0.662)
Medication	0 = no medication; 1 = medication	n_0 = 534; n_1= 64	ES_0 -3.056 (0.235); ES_1 -1.920 (0.761)
Psychological intervention	0 = no psycho-intervention component included; 1 = psycho-intervention component included	n_0 = 94; n_1= 504	ES_0 -2.388 (0.298); ES_1 -3.030 (0.225)
Teaching alternative replacement skills	0 = not teaching alternative replacement skills; 1 = teaching alternative replacement skills	n_0 = 405; n_1= 193	ES_0 -2.742 (0.249); ES_1 -3.444 (0.326)
Reward, praise, attention	0 = no reward, praise, attention; 1 = reward, praise, attention	n_0 = 290; n_1= 308	ES_0 -2.481 (0.285); ES_1 -3.384 (0.274)
Punishment	0 = no punishment; 1 = punishment	n_0 = 480; n_1= 118	ES_0 -2.878 (0.253); ES_1 -3.214 (0.485)
Use of restraints	0 = no use of restraints; 1 = use of restraints	n_0 = 562; n_1= 36	ES_0 -2.964 (0.233); ES_1 -2.913 (0.860)
Manipulating antecedent factors	0 = not manipulating antecedent factors; 1 = manipulating antecedent factors	n_0 = 366; n_1= 232	ES_0 -2.765 (0.237); ES_1 -3.243 (0.256)
Extinction	0 = no extinction; 1 = extinction	n_0 = 457; n_1= 141	ES_0 -3.033 (0.256); ES_1 -2.720

			(0.465)
Gain insight in perception of CB	0 = not gaining insight in perception of CB; 1 = gain insight in perception of CB	n_0 = 534; n_1= 64	ES_0 -2.990 (0.237); ES_1 -2.592 (0.720)
Social-contextual intervention	0 = no socio-intervention component included; 1 = socio-intervention component included	n_0 = 297; n_1= 301	ES_0 -2.623 (0.318); ES_1 -3.293 (0.313)
Informing, educating, training the environment	0 = not informing, educating, training the environment; 1 = informing, educating, training the environment	n_0 = 431; n_1= 167	ES_0 -2.755 (0.261); ES_1 -3.530 (0.429)
Adapt the environment to the participant's needs	0 = not adapting the environment to the participant's needs; 1 = adapt the environment to the participant's needs	n_0 = 361; n_1= 237	ES_0 -2.563 (0.285); ES_1 -3.575 (0.352)
Publication year	Continuous	*M* = 2005; *SD* = 3.48; range = 2000-2011	0.002 (0.064)
Study quality	Continuous	*M* = 6.803; *SD* = 1.609; range = 2-10	-0.417 (0.0315)

Notes: *M* = mean; *SD* = standard deviation; CB = challenging behavior; ES = effect size; SE = standard error.

[a] The variable 'Intellectual disability level' is included in Tables 3 and 4, but is not displayed in Table 5, since there were 32.7% missing values for this variable. We coded whether the participant had a borderline, mild, moderate, severe, or profound level of ID. For 32.7% of the included participants, the articles only mentioned that the participant had ID, but they did not describe the specific level of ID, nor could the level of ID be deduced from described test results. Since the applied analysis methods only work with participants without missing values, we decided to not include this variable in the remaining analysis.

Table 5

Parameter estimates and standard errors for the multilevel meta-analysis of the single-case and small-*n* studies.

	Model 1	Model 2	Model 3
Fixed effects			
Mean treatment effect	-2.96 (0.23) ***		-2.00 (0.47) ***
Moderator effect of			
Age		-0.06 (0.02) **	-0.04 (0.02) *
Gender		-0.27 (0.48)	
CB type – Self-injurious behavior		-0.13 (0.16)	
CB type – Stereotyped behavior		-0.30 (0.24)	
CB type – Aggression		1.18 (0.11) ***	1.01 (0.10) ***
CB type – Destructive behavior		0.32 (0.12) **	0.21 (0.12)
CB type – Disruptive behavior		0.25 (0.13)	
Diagnosis of autism spectrum disorder		-1.24 (0.36) ***	-0.95 (0.33) **
Sensory impairment		0.26 (0.90)	
Motor impairment		-0.04 (0.74)	
Communicative impairment		1.14 (0.60)	
Design		-0.36 (0.30)	
Pretreatment functional analysis		-0.16 (0.72)	
Primary treatment setting		-0.11 (0.12)	
Primary intervention agent		-0.09 (0.19)	
Duration of treatment		-0.08 (0.20)	
Intervention format		3.26 (2.49)	
Family involved in treatment		0.25 (0.91)	
Peer(s) involved in treatment		1.05 (1.93)	
Presence of data/researcher triangulation		-1.01 (1.14)	
Presence of follow-up data		-0.19 (0.58)	
Uni- vs. multicomponent intervention		0.11 (0.77)	
Biological intervention		3.43 (1.76)	
Medication		-2.24 (2.04)	
Psychological intervention		0.66 (0.59)	
Teaching alternative replacement skills		-0.44 (0.51)	
Reward, praise, attention		-0.66 (0.53)	
Punishment		-0.19 (0.64)	
Use of restraints		-0.01 (0.96)	
Manipulating antecedent factors		-1.52 (0.54) **	-0.50 (0.21) *
Extinction		0.03 (0.62)	
Gain insight in perception of CB		0.02 (1.26)	
Social-contextual intervention		1.86 (1.28)	
Informing, educating, training the environment		-2.30 (0.86) **	-0.92 (0.53)
Adapt the environment to the participant's needs		-1.77 (1.06)	
Publication year		-0.04 (0.07)	
Study quality		0.09 (0.21)	
Variance of effect			
Between studies [a]	3.32 (0.91) ***	2.51 (1.18) *	3.33 (0.97) **
Between participants [b]	20.57 (1.37) ***	21.90 (1.62) ***	20.40 (1.42) ***
Residual variance	1.00 (0.01) ***	0.99 (0.01) ***	0.99 (0.01) ***

Note: * = statistically significant effect: $p < .05$; ** = $p < .01$; *** = $p < .001$

[a] Variance of effect between studies for Model 1 for the database without the most extreme outlying participant: 4.30 (1.02) ***; for the database without the six most extreme outlying participants: 3.72 (0.54) ***

[b] Variance of effect between participants for Model 1 for the database without the most extreme outlying participant: 11.76 (0.95) ***; for the database without the six most extreme outlying participants: 3.48 (0.30) ***

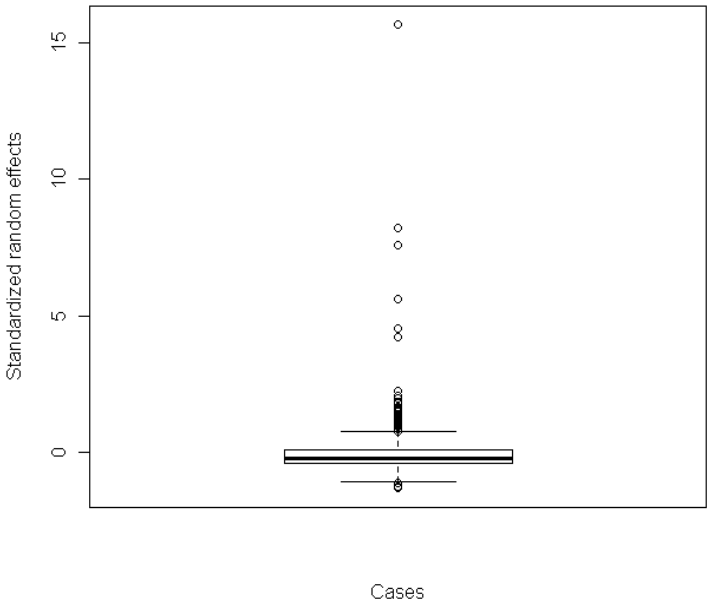

Figure 1: Box and whisker diagram of the standardized random effects for the individual participants.

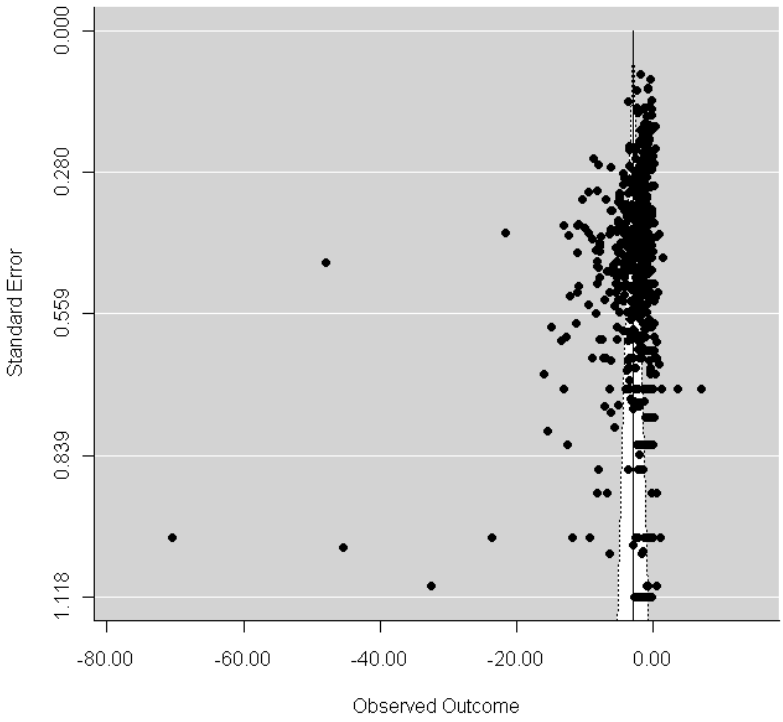

Figure 2: Funnel plot for publication bias: Each black dot represents the observed outcome for one participant.